Spring 5-12-2011

DNA Sequence Variation in the Wingless Gene Product in Buckeye Butterflies (Genus Junonia)

Bonnie McCullagh
Western Kentucky University, bonnie.mccullagh686@topper.wku.edu

ABSTRACT

Wingless is a highly conserved gene important to cell determination in development. In *Drosophila*, the wingless gene product has been identified as responsible for wing patterning. In *Bicyclus anynana* and *Junonia coenia*, wingless gene product is expressed in a fashion that suggests that it is involved in butterfly wing color pattern development. The wingless gene product has been implicated as a potential focal signal for patterning the eyespot of *Junonia* butterflies. I have shown that extensive DNA sequence variation (26.04% of the sequenced region) exists in 402 bp of *wingless* coding sequence among 338 specimens of *Junonia* from Florida, Texas, Kentucky, California, and Argentina, representing 6 nominal species. Much of the identified variation is synonymous, but it alters codon usage and therefore has the potential to affect the amount of gene product produced. A common haplotype that uses unfavored codons has been identified and this might account for smaller eyespots. Positions of non-synonymous variation have been also been identified among the samples studied which may affect the behavior of the *wingless* gene product. The variation in *wingless* sequence was also used to examine geographical and nominal species population structure among different *Junonia* populations.

Keywords: DNA Sequencing, Butterfly Eyespots, Polymerase Chain Reaction, *Junonia*, Coding sequence, Population structure

ACKNOWLEDGMENTS

I would like to thank Dr. Jeffrey Marcus for his invaluable assistance and guidance with this project. I would also like to thank my committee members for their help in this process. I thank Dr. Rudi Mattoni for providing *Junonia* specimens from Argentina and Bob Dowell for specimens from California. Thanks also go to Tanja Borchers and Colin Rumbolt for their assistance in the lab. I would like to thank my friends and family for their support during my undergraduate career. Sarah Crites, Jenny Kiefer, and Sarah Randall, I cannot thank you all enough for coming to my presentations and listening to me "talk science." I would also like to thank Joanne Seiff for always being my cheerleader and full of support. This research was supported by an Honors Development Grant from the WKU Honors College, a Biology Summer Undergraduate Research Experience Fellowship from the Ogden College of Science and Engineering, and an Undergraduate Research Program grant from the Kentucky Academy of Sciences.

VITA

April 20, 1990……..……..…Born – Midland, Texas

2007……...…Christian County High School, Hopkinsville, Kentucky

2008-2011……Undergraduate Researcher in the lab of Dr. Jeffrey Marcus

2009……..Presentation at the Kentucky Academy of Sciences

2009…….……..Summer Research Experience at Harvard Medical School

2010…….Presentation at The 6th International Conference on the Biology of Butterflies

2010……..Presentation at the Kentucky Academy of Sciences

2010…….……..Summer Research Experience at the University of Manitoba

FIELDS OF STUDY

Major Field: Biology

Major Field: History

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1

INTRODUCTION

Wingless is a highly conserved gene that encodes a secreted ligand with important roles in cell determination during development. In *Drosophila melanogaster*, *wingless* is involved in cell-cell communication that mediates signal transduction. *Wingless* controls the segmentation pattern of embryos along with patterning the imaginal disk (GILBERT 2003; LINDSLEY and ZIMM 1992). Wingless is essential for establishing the marginal boundary of the developing wing in *Drosophila,* and is named for the mutant phenotype observed in hypomorphic (partial loss of function) *Drosophila* mutants (LINDSLEY and ZIMM 1992). The well-characterized nature of the *wingless* gene in *D. melanogaster* allows for inferences about the role of *wingless* in other organisms.

In vertebrates, *wingless* (known as *Wnt*) comprises a family of 15 different genes. As observed in *Drosophila*, this family of proteins is involved in cell development. Wnt4 is involved in female sex determination along with kidney development. Wnt1 has a role in the formation of muscle cells along with urogenital development. Other Wnt proteins play roles as varied as the specification of neural crest to the patterning of the developing vertebrate limb bud (GILBERT 2003). The *Wnt* family of genes is recognizable across diverse taxa (from hydra to humans) and evolves relatively slowly due to its important roles in fundamental developmental processes (HOBMAYER *et al.* 2000).

In addition to its roles in embryonic development and its role in defining the wing

margin, *wingless* also induces the development of color patterns in adult wings during metamorphosis. In *Drosophila* species that have pigment patterns on their wings, wingless expression in the imaginal disk determines the location of pigmented areas (WERNER *et al.* 2010). In butterflies, *wingless* has been implicated as possibly having roles in the development of two different types of color patterns. In the buckeye butterfly, *Junonia coenia*, *wingless* mRNA is expressed in the wing discs during the development of the pair of orange bands that compose the central symmetry system in this genus (CARROLL *et al.* 1994). In *Bicyclus anynana*, the wingless protein is expressed in the eyespot focus (MONTEIRO 2008). *Wingless* expression during the development of these color patterns in butterflies suggests that it may have a functional role, and *wingless* is a good candidate for being the focal signal that patterns eyespots (MARCUS 2005; NIJHOUT 1991).

Wingless is also an important tool in molecular phylogenetics research, particularly in butterflies, where it is frequently used to infer evolutionary relationships (BROWER 2000; BROWER and DESALLE 1998). *Wingless* was one of several genes used to establish that the buckeye butterflies are actually members of the genus *Junonia* rather than in the genus *Precis* (KODANDARAMAIAH and WAHLBERG 2007; WAHLBERG et al. 2005). The combination of an understanding of its molecular function, with its utility in understanding evolutionary relationships makes *wingless* a unique candidate gene by which phylogenetic relationships among butterflies can be connected to the evolution of wing color pattern phenotypes. I have sequenced and analyzed a 402 bp segment of the *wingless* coding sequence from buckeye butterflies (genus *Junonia*) collected in Florida, Texas, Kentucky, California, and Argentina. This 402 bp segment of *wingless* coding

sequence contains approximately half of the coding sequence and is located in the last exon in *wingless* (Figure 1). The specimens that I studied belong to several different named species including the common buckeye (*J. coenia coenia*), the mangrove buckeye (*J. evarete zonalis)*, and the tropical buckeye *(J. genoveva genoveva)* in Florida*;* the common buckeye (*J. coenia coenia*) and the dark buckeye (*J. sp. nigrosuffusa*) in Texas; the common buckeye (*J. coenia coenia)* in Kentucky; the common buckeye *(J. coenia grisea)* in California (Figure 2); and the mangrove buckeye (*J. evarete flirtea*), and the tropical buckeye *(J. genoveva hilaris*) in Argentina (Figure 3). Species designations were made according to Turner and Parnell (TURNER and PARNELL 1985). There is extensive phenotypic variation in eyespot size and structure among and between the various forms of *Junonia* in the new world (CECH and TUDOR 2005; GLASSBERG 2007; GLASSBERG *et al.* 2000), and previous work by the Marcus laboratory (unpublished) has shown that there is evidence of hybridization and mitochondrial "capture" between the different forms of this genus (Figure 4).

CHAPTER 2

METHODS AND MATERIALS

Junonia butterflies were captured from natural habitats in Florida, Texas, Kentucky, California, and Argentina from 2004 to 2010 using a hand-held butterfly net. Global Position System (GPS) coordinates were recorded for each specimen collected using a handheld Garmin GPS 12XL. Butterflies were stored alive in glassine envelopes (BioQuip) at 4°C while being transported back to the laboratory, at which point they were frozen and maintained at -20°C.

DNA was isolated from the legs of frozen *Junonia* buckeye specimens using the Qiagen DNEasy kit. Polymerase Chain Reaction (PCR) was performed using degenerate PCR primers for the *wingless* gene Lepwg1 and Lepwg2 (BROWER and DESALLE 1998). The reaction volumes were 25 μL and the reaction conditions used to amplify *wingless* were 94°C for 5 minutes; 40 cycles of 94°C for 1 minute, of 46°C for 1 minute, 72°C for 2 minutes; and a 10 minute final extension at 72°C before being placed on a 4°C hold using a BioRad MyCycler Thermocycler. PCR products were loaded onto a 1% agarose gel in TAE buffer and gels were run for 1 hour at 77 V to determine if successful amplification had occurred. The gels were then stained in an Ethidium Bromide solution and visualized using UV illumination. PCR product size (450 bp) was confirmed by comparison to a 1 KB ladder (New England Biolabs) (Figure 5).

The PCR products were sequenced using Sanger dideoxy sequencing. Sequencing reactions were performed in 11 uL reactions using BigDye V3.1 Dye Termination

sequencing chemistry (Applied Biosystems). The sequencing reactions were then placed in a BioRad MyCycler Thermocycler and subjected to the following conditions: 96°C for 2 minutes; 25 cycles at 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes; ending with a final hold at10°C.

In order to purify the sequencing reactions, the 11 μL sequencing reactions were transferred by pipetting to 1.5 mL microcentrifuge tubes. Forty μL of 75% isopropanol was added to each sequencing microcentrifuge tube. Each tube was inverted and vortexed to mix. The tubes then incubated at room temperature for 15 minutes. After incubation, the tubes were placed in a Sorvall Legend microcentrifuge and spun at maximum speed (13.2 x 1000 RPM) for 20 minutes at room temperature. The supernatant was removed by pipetting and an additional 250 μL of 75% isopropanol was added. Tubes were inverted and vortexed and then spun at maximum speed for 5 minutes. The supernatant was removed by pipetting 125 μL volumes twice. The pellets were dried in a speed vacuum for 15 minutes without heat. The samples were then loaded onto an ABI 3130 Sequencer.

Sequence data was obtained from an ABI 3130 Sequencer. Forward and Reverse sequences for each individual were assembled into contigs using Sequencher software v. 4.6 (GeneCodes), followed by manual spot-checking for discrepancies. Primer sequences were trimmed from the assembled contigs, yielding 402 bp sequences for further analysis. Sequences were then checked for Single Nucleotide Polymorphisms (SNPs). The SNPs were identified by manual examination and comparison of the 402 bp sequence chromatograms using Sequencher. All of the assembled contigs were exported in FASTA format and aligned using Clustal W 2.0.12 (THOMPSON *et al.* 1994). The positions of polymorphisms in the sequences were determined by alignments in Clustal W. Initial

alignments in Clustal W were run to verify the existence of SNPs in smaller portions of the data set. Alignments of the entire data set were then run in Clustal W to verify the location of SNPs. After the multiple alignments in Clustal W, SNPs were double-checked by examination of the sequence chromatograms in Sequencher.

Within the 402 bp sequence, 88 sequence positions showed polymorphism among 338 *Junonia* specimens. The nucleotides at the 88 positions were analyzed using PHASE v2.1 to obtain haplotype frequencies and predicted haplotype pairs for each individual (STEPHENS *et al.* 2001). The haplotype pairs predicted by PHASE v2.1 relay which 2 haplotypes from the population are most likely held by an individual. The 4 most frequent alleles with a frequency greater than 0.04 were identified for further analysis. Percentages of codon usage were generated by using the following in conjunction: Graphical Codon Usage Analyser 2.0 (FUHRMANN *et al*. 2004) and a codon usage table from the Codon Usage Database (NAKAMURA 2007) and based on codon usage in *invected, extra sex combs, ASH1, patched, distalless, ultrabithorax*, and *fringe* mRNA coding sequences. Codon usage differences between alleles were recorded. Preferred and unfavored codon substitutions were recorded for each allele.

The predicted most likely haplotype pairs for each specimen (as listed in the Best Pairs Summary) from PHASE analysis were utilized to determine which individuals in the sample contained the most frequent alleles. The allele frequency data was then used to analyze relationships between species and geographic locations corresponding to individuals possessing these most frequent alleles.

Using the Best Pairs data from PHASE, the program GENEPOP 4.0.10 (RAYMOND and ROUSSET 1995; ROUSSET 2008) was utilized to test for compliance with

Hardy-Weinberg equilibrium. The importance of heterozygotes to populations and subpopulations was determined by these tests of allelic variation in GENEPOP 4.0.10. The populations were then separated on the basis of geographic location and analyzed for genic differences in population structure. Subpopulations were further subdivided by species within each location to analyze genic differences among the different species.

Finally, from among the 338 *Junonia* specimens, selected individuals were pinned and spread (WINTER 2000), to allow examination of variation in wing phenotypes (Figure 3). Specimens were digitally photographed for future analysis.

CHAPTER 3

RESULTS

A total of 338 *Junonia* specimens from Florida, Texas, Kentucky, California, and Argentina were successfully surveyed at the *wingless* locus. The 402 bp DNA sequence region revealed substantial variation including 88 variable positions among the 402 bp region and within the 338 samples. The identification of 88 variable positions is significant due to the highly conserved nature of the *wingless* gene. Any samples that failed to produce the 402 bp DNA sequence failed to amplify in whole, likely due to poor DNA preservation of the sample. Of the 88 sites identified in the dataset, 65 positions showed synonymous DNA variation (73.86% of the synonymous sites). Synonymous sites comprised 26.04% of all positions in the 402 bp DNA sequence region. An additional 23 positions were found to show non-synonymous variation (26.14% of the variable sites). The non-synonymous sites comprised 6.80% of all sites in the 402 bp DNA region.

Among the non-synonymous variation detected, 15 of the 23 changes were non-conservative amino acid changes, and 1 was a nonsense mutation that created a stop codon near the carboxy terminus of the protein (Table 1). The remaining 7 conservative non-synonymous substitutions are unlikely to result in major changes in *wingless* protein function. Of the 23 non-synonymous substitutions, only 7 of the amino acid changes were found in more than 1 individual. The amino acid changes that occur in more than one individual represent a larger portion of the sample as the amino acid changes occurring once confer only a frequency of 0.3%. Therefore, the amino acid changes in multiple

individuals further confirms the presence of these variable positions. Only 3 non-conservative amino acid changes were found in multiple individuals. Although substantial non-synonymous variation exists in the wingless gene, much of the non-synonymous variation is conservative in nature and therefore unlikely to have a measureable effect on the wingless gene. The non-conservative and non-synonymous variation present is almost entirely limited to single representatives in the sampled population for each polymorphism and may represent rare deleterious alleles that are not viable when homozygous or they may represent PCR artifacts or sequencing errors.

The genetic variation found in the *wingless* gene in *Junonia* was analyzed further using PHASE software. PHASE software was used to analyze allelic frequencies in a population and construct haplotypes present in the population. According to PHASE analysis allele frequencies, as many as 731 different alleles could be present in the population. This number of alleles is generated by analysis of the genotypes of each individual in the dataset. However, the Best Pairs Summary from PHASE showing the most probable alleles for each individual suggested that only 219 alleles were present in the pool of 338 specimens. The 4 most common alleles (with predicted allele frequencies of 4.60% to 10.45%) were selected for further analysis. Alleles 45 (4.60%), 177 (4.99%), 89 (7.84%), and 72 (10.45%) showed the highest frequencies among the sampled population. Using the Graphical Codon Usage Analyzer (FUHRMANN *et al*. 2004), allelic differences in codon usage were determined. As the most common allele, allele 72 was used as the standard for comparison. Allele 89 had a very small (1 codon) negative net change in preferred codon usage compared to allele 72. Allele 45 had a 2 codon negative net change in preferred codon usage compared to the standard. However, allele 177 was

found to encode 1 additional favored codon and 5 additional unfavored codons, when compared to allele 72, for a negative net change of 4 codons. Individuals carrying allele 177 might therefore be expected to produce reduced amounts of wingless protein. If wingless is the focal signal for eyespot patterning, smaller eyespots might be expected in individuals that carry this allele. According to Best Pair data, 2 individuals in the dataset are homozygous for this allele and may therefore be expected to have smaller eyespots.

Using GENEPOP for further analysis of population structure, the population was checked for compliance with Hardy-Weinberg equilibrium using the Markov chain method (RAYMOND and ROUSSET 1995; ROUSSET 2008). Hardy-Weinberg equilibrium analysis is useful for determining the structure of populations and subpopulations in the dataset. The GENEPOP analysis failed to reject the null hypothesis of Hardy-Weinberg equilibrium when analyzed by population location (Texas, Florida, Kentucky, California, and Argentina) with the exception of Florida (P-value=0.0004). Additionally, Florida populations were found to have a deficit of heterozygotes (P=0.0042) whereas the other populations did not show a deficit of heterozygotes. Florida populations, based on genotypes of individuals, were expected to contain more heterozygotes. GENEPOP analysis was also conducted with populations divided by nominal species. *J. coenia* and *J. evarete* rejected Hardy-Weinberg equilibrium. The two species were also found to have a deficit of heterozygotes (P=0.0089 and 0.0121, respectively) meaning that these species contain fewer than the expected number of heterozygotes.

Using GENEPOP, genic differentiation analyses were conducted to reveal structure between subpopulations on the basis of species and location (Table 2). When compared using pairwise genic differentiation, Florida *Junonia* species showed

substantial variation between each named species. Significant genic differentiation was observed between Texas mangrove and Florida common species. Surprisingly, there is substantial genic variation between Texas and Florida *J. coenia*, indicating geographic separation. Texas species also exhibit population structure providing support for the presence of 3 nominal species in Texas. Kentucky *J. coenia* showed similarity to most subpopulations with the exception of Texas *J. coenia* and Texas *J. sp. nigrosuffusa*. This may indicate geographic distinction between these populations. Argentina species show genic differentiation with most other comparisons indicating geographic separation. However, Argentina species are not significantly different when compared to Florida mangrove and tropical species suggesting genetic similarities between these populations. California species had great dissimilarity with all species except Texas mangroves. These analyses suggest that geography is not always a barrier to gene flow and that interspecies gene transfer can be fluid (Figure 6).

Fst analyses were also conducted using GENEPOP to examine what proportion of population structure occurs within individuals in subpopulations versus what proportion of population structure occurs between subpopulations of the total population. If population structure is present, it indicates that differences among subpopulations are larger than differences between individuals in the subpopulations. When the Fst value exceeds 0.01, the amount of structure is considered significant. These analyses revealed the amount of genetic differentiation between populations. Florida *J. genoveva* and *J. coenia* had little genetic differentiation between them yet were different from Florida *J. evarete* species. Florida, Kentucky, and Texas *J. coenia* populations have little genetic differentiation between them, but are distinct from *J. coenia* populations in California.

Argentina *J. evarete* and *J. genoveva* have very little genetic differentiation in their populations, potentially due to hybridization between these species in this region (Figure 6).

When the sampling locations of specimens carrying the most common *wingless* alleles were analyzed, several patterns appeared. Some common alleles span multiple geographic localities and cross nominal species boundaries. Allele 45 was found in specimens from Argentina, Florida, and Texas. However, the allele did not appear in all nominal species from each locality. Allele 45 appeared in all Florida species, Texas *J. coenia* and *J. sp. nigrosuffusa*, and Argentina *J. genoveva hilaris*. Another frequent allele, allele 177, has interesting structure because it occurs in *J. coenia* species from California, Florida, and Texas. Additionally, allele 177 appears in Florida *J. evarete* and *J. genoveva* species, meaning that allele 177 does not appear to be restricted by species. Samples carrying allele 89 were found in all Florida species, Texas *J. coenia*, Kentucky *J. coenia*, and Argentina *J. genoveva hilaris.* The most abundant allele, Allele 72, was found in all 3 Florida buckeye species and was also found in 1 Argentina sample, *J. evarete flirtea*.

CHAPTER 4

DISCUSSION

Wingless is an essential gene for embryonic development that is conserved across the metazoans (HOBMAYER *et al.* 2000). It is therefore somewhat surprising that 26.04% of the 402 *wingless* nucleotide positions surveyed in *Junonia* are variable, and that an estimated 219 different alleles, according to PHASE analysis, were detected in a pool of just 338 individuals. Most of this variation was synonymous, but some of the changes in common alleles produce alterations in codon usage that could affect the translation efficiency of *wingless* mRNAs. Rare alleles with non-synonymous sequence variation were also detected in the population. The substantial sequence variation observed in a gene like *wingless* that is essential for normal development suggests that at least some of this variation may have functional consequences on the gene.

The amount of variation in *wingless* indicates that heterozygous sequence variation may be advantageous in *Junonia*. A hypothesis of heterozygote advantage generally predicts an overabundance of heterozygotes (FORD 1975), but in the *Junonia* data set, there is no significant excess of heterozygotes from what would be predicted by Hardy-Weinberg. However, there are so many alleles segregating (minimum 219) at the *wingless* locus in *Junonia*, that homozygotes are extremely rare (5 out of 338 individuals sampled). The sample size necessary to detect an excess of heterozygotes, according to Hardy-Weinberg equilibrium, may therefore be much larger than the sample size in these

experiments. Additional sampling from these populations may allow for a better test of heterozygote advantage in the future.

An alternative possibility is that *wingless,* in spite of being an evolutionarily conserved gene, may be evolving quickly within *Junonia* populations. However, if this is the case, the new alleles are not replacing old ones, but rather coexisting with them for long periods of evolutionary time, since common alleles are shared between Florida, Texas, California, and Kentucky, as well as between North and South America. Perhaps a frequency dependent selection regime, where common alleles are at a selective disadvantage, might explain the patterns of diversity observed in the dataset.

On the basis of codon usage, individuals carrying alleles with an excess of unfavored codons would be expected to have smaller eyespots than individuals carrying alternate alleles. Within the variation sampled, one allele (177) with multiple unfavored codons has the potential to produce this type of phenotypic effect. The allele was found in *J. coenia* from Florida, Texas, and California as well as *J. evarete* and *J. genoveva* from Florida. Comparing eyespots and orange bands in individuals with and without this allele from these populations will further our understanding of the phenotypic consequences of this allelic variation. Because this allele has unfavored codons, smaller eyespots would be expected. However, there may be multiple loci that contribute to producing eyespot variation in *Junonia. J. evarete* typically has smaller eyespots but allele 177 does not appear in samples from every geographic location with *J. evarete.* It will be very interesting to analyze the phenotypic variation in eyespot size among these individuals in conjunction with the molecular variation that we have observed thus far.

Regardless of its phenotypic consequences, the variation in *wingless* sequence is

also useful for studying population structure in *Junonia*, especially as the findings from this nuclear marker can be compared with previously observed patterns in mitochondrial haplotypes (Figure 4). Using *wingless* sequence data, it is possible to separate populations on the basis of location and nominal species (Figure 6). While Cytochrome Oxidase I (COI) mitochondrial haplotypes showed evidence of mitochondrial capture due to hybridization in *Junonia,* the analysis of population structure based on wingless sequences shows that in Florida one of the nominal species *(J. evarete)* remains genetically distinct from Florida *J. coenia* and *J. genoveva.* The three nominal forms in Texas have some genetic structure between them. Texas *J. evarete* and *J. sp. nigrosuffusa* have genetic structure in these two forms that does not appear when comparing *J. evarete* to *J. coenia* or *J. coenia* to *J. sp. nigrosuffusa.* Interestingly, common buckeyes *(J. coenia)* in Texas, Kentucky, and Florida are not distinct from each other, suggesting that in at least in some forms of *Junonia*, there is substantial population connectivity over large geographic distances. The two *Junonia* forms in Argentina are similar genetically despite having different species designations. *J. evarete flirtea* has very little genetic differentiation when compared to Kentucky and Florida *J. coenia* populations, a further indication of genetic connectivity over large distances. Argentinean *J. genoveva hilaris*, however, has little differentiation from Texas *J. sp. nigrosuffusa*. The California *J. coenia* species shows moderate genetic differentiation when compared to the same species in different localities and different *Junonia* species. This indicates that the California *J. coenia* population may be geographically isolated in a way that the other populations are not.

Previous attempts to classify Lepidoptera species have focused on the use of the

mitochondrial marker COI because of its slow evolution in many species. Much has been made about the utility of the COI barcodes for identifying and understanding relationships among Lepidoptera species (HEBERT et al. 2004; JANZEN et al. 2005; MARCUS et al. 2010), but because of extensive hybridization and a history of mitochondrial capture, COI is a very poor marker for making species determinations in *Junonia* (Figure 4). Nuclear wingless sequences may be far more valuable for assigning specimens to a particular species than COI, especially since species determinations of New World *Junonia* specimens are also extremely difficult on the basis of visible phenotypes (NEILD 2008).

LITERATURE CITED

BROWER, A. V. Z., 2000 Phylogenetic relationships among the Nymphalidae (Lepidoptera) inferred from partial sequences of the wingless gene. Proc. R. Soc. B. 267: 1201-1211.

BROWER, A. V. Z., and R. DESALLE, 1998 Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: The utility of wingless as a source of characters for phylogenetic inference. Ins. Mol. Biol. 7: 73-82.

CARROLL, S. B., J. GATES, D. N. KEYS, S. W. PADDOCK, G. E. F. PANGANIBAN et al., 1994 Pattern formation and eyespot determination in butterfly wings. Science 265: 109-114.

CECH, R., and G. TUDOR, 2005 Butterflies of the East Coast. Princeton University Press, Princeton, NJ.

FORD, E. B., 1975 Ecological Genetics. Chapman and Hall, London.

FUHRMANN, M., A. HAUSHERR, L. FERBITZ, T. SCHÖDL, M. HEITZER et al., 2004 Monitoring dynamic expression of nuclear genes in Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. Plant Mol. Biol. 55: 869-881.

GILBERT, S. F., 2003 Developmental Biology. Sinauer Associates, Inc., Sunderland, MA.

GLASSBERG, J., 2007 A Swift Guide to the Butterflies of Mexico & Central America. Sunstreak, Morristown, NJ.

GLASSBERG, J., M. C. MINNO and J. V. CALHOUN, 2000 Butterflies through binoculars: A field, finding, and gardening guide to butterflies in Florida. Oxford, New York.

HEBERT, P. D. N., E. H. PENTON, J. M. BURNS, D. H. JANZEN and W. HALLWACHS, 2004 Ten species in one: DNA barcoding reveals cryptic specie

sin the neotropical skipper butterfly Astraptes fulgerator. Proc. Nat. Acad. Sci. USA 101: 14812-14817.
HOBMAYER, B., F. RENTZSCH, K. KUHN, C. M. HAPPEL, C. C. VON LAUE et al., 2000 WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. Nature 407: 186-189.
JANZEN, D. H., M. HAJIBABAEI, J. M. BURNS, W. HALLWACHS, E. REMIGIO et al., 2005 Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Phil. Trans. Roy. Soc. B 360: 1835-1845. (doi:1810.1098/rstb.2005.1715).
KODANDARAMAIAH, U., and N. WAHLBERG, 2007 Out-of-Africa origin and disperal-mediated diversification of the butterfly genus Junonia (Nymphalidae: Nymphalinae). J. Evol. Biol. 20: 2181-2191.
LINDSLEY, D. L., and G. G. ZIMM, 1992 The Genome of Drosophila melanogaster. Academic Press, San Diego, CA.
MARCUS, J. M., 2005 Jumping genes and AFLP maps: Transforming Lepidopteran color pattern genetics. Evol. Dev. 7: 108-114.
MARCUS, J. M., T. M. HUGHES, D. M. MCELROY and R. E. WYATT, 2010 Engaging First Year Undergraduates in Hands-On Research Experiences: The Upper Green River Barcode of Life Project. J. Coll. Sci. Teach. 39.
MONTEIRO, A., 2008 Alternative models for the evolution of eyespots and of serial homology on lepidopteran wings. BioEssays 30: 358-366.
NAKAMURA, Y., 2007 Codon Usage Database, Department of Plant Genome Research, Kazusa DNA Research Institute, http://www.kazusa.or.jp/codon/.
NEILD, A. F. E., 2008 The Butterflies of Venezuela, Part 2: Nymphalidae II (Acraeinae, Libytheinae, Nymphalinae, Ithomiinae, Morphinae). Meridian, London.
NIJHOUT, H. F., 1991 The development and evolution of butterfly wing patterns. Smithsonian Institution Press, Washington.
RAYMOND, M., and F. ROUSSET, 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenism. J. Hered. 86: 248-249.
ROUSSET, F., 2008 Genepop'007: a complete reimplementation of the Genepop software
for Windows and Linux. Mol. Ecol. Resources 8: 103-106.
STEPHENS, M., N. SMITH and P. DONNELLY, 2001 A new statistical method for haplotype
reconstruction from population data. . Am. J. Hum. Genet. 68: 978-989.
THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.
TURNER, T. W., and J. R. PARNELL, 1985 The identification of two species of Junonia Hubner (Lepidoptera: Nymphalidae): J. evarete and J. genoveva in Jamaica. J. Res. Lep. 24: 142-153.
WAHLBERG, N., A. V. Z. BROWER and S. NYLIN, 2005 Phylogenetic relationships and historical biogeography of tribes and genera in the subfamily Nymphalinae (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. 86: 227-251.
WERNER, T., S. KOSHIKAWA, T. M. WILLIAMS and B. J. CARROLL, 2010

Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464: 1143-1148.
WINTER, W. D., 2000 Basic Techniques for Observing and Studying Moths and Butterflies. The Lepidopterists' Society, Los Angeles, CA.

Tables

Table 1. Non-synonymous substitutions in the sampled *wingless* coding sequence

Sequence Position	Common Haplotype (frequency) (a)[1]	Rare Haplotype (frequency) (b)	Nature of Change
44	Arginine (675)	Histidine (1)	conservative
88	Arginine (674)	Tryptophan (2)	non-conservative
100	Proline (675)	Threonine (1)	non-conservative
103	Asparagine (675)	Tyrosine (1)	non-conservative
106	Threonine (672)	Serine (4, 1 homozygote)	conservative
110	Glycine (675)	Glutamic acid (1)	non-conservative
113	Isoleucine (675)	Threonine (1)	non-conservative
115	Glutamic acid (675)	Glutamine (1)	conservative
122	Proline (675)	Histidine (1)	non-conservative
125	Valine (675)	Glycine (1)	non-conservative
139	Valine (674)	Isoleucine (2)	conservative
154	Histidine (675)	Phenylalanine (1)	non-conservative
190	Proline (675)	Alanine (1)	non-conservative
205	Histidine (675)	Tyrosine (1)	non-conservative
292	Serine (674)	Threonine (2)	conservative
310	Asparagine (675)	Histidine (1)	non-conservative
317	Alanine (675)	Glutamic acid (1)	non-conservative
340	Aspartic acid (674)	Asparagine (2)	non-conservative
346	Glycine (675)	Alanine (1)	non-conservative
371	Threonine (675)	Serine (1)	conservative
391	Glutamic acid (671)	Lysine (5)	conservative
394	Arginine (673)	Glycine (2)	non-conservative
		Stop (1)	nonsense

[1] 338 individuals sampled were all diploid resulting in 676 alleles. Unless otherwise noted, rare alleles were heterozygous. (a) refers to the amino acid carried by most individuals at this position. (b) refers to the non-synonymous change in amino acids at this position. The number in parentheses refers to the number of alleles coding for the listed amino acid.

Table 2: Pairwise P-values for GENEPOP analysis of population structure in *Junonia.*[2]

Population pair	P-Value	S.E.	Switches
FL Mangrove & FL Common	0.00000	0.00000	2779413
FL Tropical & FL Common	0.00228	0.00014	2746263
FL Mangrove & FL Tropical	0.00000	0.00000	2312841
TX Mangrove & FL Common	0.14314	0.00168	157458
TX Mangrove & FL Tropical	0.02107	0.00060	106246
TX Mangrove & FL Tropical	0.08611	0.00090	307377
TX Common & FL Common	0.14912	0.00157	2713145
TX Common & FL Tropical	0.00000	0.00000	2245746
TX Common & FL Tropical	0.00000	0.00000	2836304
TX Common & TX Mangrove	0.42475	0.00239	225114
TX Dark & FL Common	0.06296	0.00107	1589403
TX Dark & FL Tropical	0.00000	0.00000	1104729
TX Dark & FL Tropical	0.00011	0.00002	2381633
TX Dark & TX Mangrove	0.53596	0.00158	652113
TX Dark & TX Common	0.64381	0.00209	1960530
KY Common & FL Common	0.00518	0.00026	921301
KY Common & FL Tropical	0.00000	0.00000	625934
KY Common & FL Tropical	0.00009	0.00002	1517111
KY Common & TX Mangrove	No information		
KY Common & TX Common	0.56657	0.00219	1257648
KY Common & TX Dark	0.10493	0.00095	2314540
ARG Tropical & FL Common	0.02917	0.00067	1517269
ARG Tropical & FL Tropical	0.00000	0.00000	1053356
ARG Tropical & FL Tropical	0.00000	0.00000	2280040
ARG Tropical & TX Mangrove	0.45780	0.00130	704937
ARG Tropical & TX Common	0.32943	0.00207	1892796
ARG Tropical & TX Dark	0.34485	0.00128	2727880
ARG Tropical & KY Common	0.18358	0.00098	2406448
ARG Mangrove & FL Common	0.39489	0.00256	1160952
ARG Mangrove & FL Mangrove	0.00076	0.00010	809162
ARG Mangrove & FL Tropical	0.00436	0.00015	1872758
ARG Mangrove & TX Mangrove	0.64298	0.00109	952455
ARG Mangrove & TX Common	0.35438	0.00214	1494463
ARG Mangrove & TX Dark	0.18912	0.00107	2552949
ARG Mangrove & KY Common	0.61398	0.00114	2585785
ARG Mangrove & ARG Tropical	0.18697	0.00089	2593195
CA Common & FL Common	0.00013	0.00003	1519219

[2] Specimens were separated into subpopulations on the basis of geography and nominal species.

CA Common & FL Tropical	0.00000	0.00000	1096452
CA Common & FL Tropical	0.00000	0.00000	2356685
CA Common & TX Mangrove	0.12783	0.00070	777534
CA Common & TX Common	0.00170	0.00012	1878514
CA Common & TX Dark	0.00003	0.00001	2766144
CA Common & KY Common	0.00025	0.00003	2484440
CA Common & ARG Tropical	0.00000	0.00000	2765777
CA Common & ARG Mangrove	0.00011	0.00002	2703409

Figure Captions.

Figure 1. Structure of *wingless* in *Drosophila melanogaster* with LepWg1 and LepWg2 primers annotated. Exons are represented by boxes with 5 exons in total. The introns between the exons are represented with their size in kb or bp. *Wingless* in *D. melanogaster* has an extra intron between Exon 4 and Exon 5 that separates the primer binding sites. This intron is not present in *Junonia*. White sections indicate non-coding regions of the gene; these non-coding regions have not been well studied. Black sections indicate coding regions of *wingless* that result in *wingless* gene product.

Figure 2. Photographs of the different North American *Junonia* species sampled. The 4 photographs show (from top left and moving clockwise) *Junonia coenia*, *J. sp. nigrosuffusa*, *J. evarete,* and *J. genoveva,* respectively, in the wild.

Figure 3. Photographs of the different South American *Junonia* species sampled. The 2 photographs show (from left to right) pinned specimens of *J. genoveva hilaris* and *J. evarete flirtea,* respectively.

Figure 4. Neighbor-Joining tree of partial mitochondrial cytochrome oxidase I (COI) sequences from *Junonia*. A 650 bp sequence fragment was the focus of this analysis, which included published sequences from Genbank and unpublished sequences generated by the Marcus laboratory. The analysis supports monophyly of the genus *Junonia*, and of New World representatives of the genus. Species COI haplotypes are generally not monophyletic with respect to each other, probably due to interspecific hybridization events. It appears that *Junonia evarete* has captured mitochondria from *J. coenia* on several occasions and possibly also from *J. genoveva*.

Figure 5. Agarose gel (1%) in TAE buffer after 1 hour of electrophoresis at 77 V. Successful amplification of wingless products is visible (453 bp) next to a 1kb ladder (New England Biolabs). The PCR product is 450 bp due to a LepWg1 primer size of 26 bp and a LepWg2 primer size of 25 bp, and amplified intervening sequence of 402 bp. Samples shown are *Junonia* specimens from Florida.

Figure 6. Representation of the genetic population structure of the sampled *Junonia* populations. The most likely allele pairs for each individual were determined from the Best Pairs Summary from PHASE V.2.1.1. These allele assignments were used to obtain genic differences from GENEPOP based on nominal species and location. In Florida, the 3 *Junonia* species were assigned to 3 distinct subpopulations. In Texas, there is some population structure among *Junonia evarete* and *J. sp. nigrosuffusa*. No significant population structure was observed between *J. coenia* from Kentucky, *J. coenia* from Texas, *J. coenia* from Florida, *J. evarete* from Texas, or *J. evarete flirtea* from Argentina, suggesting that there is population connectivity across these vast geographic distances. California *J. coenia* and Florida *J. evarete* exhibited population structure that separates them from other subpopulations on the basis of species and geographic location.

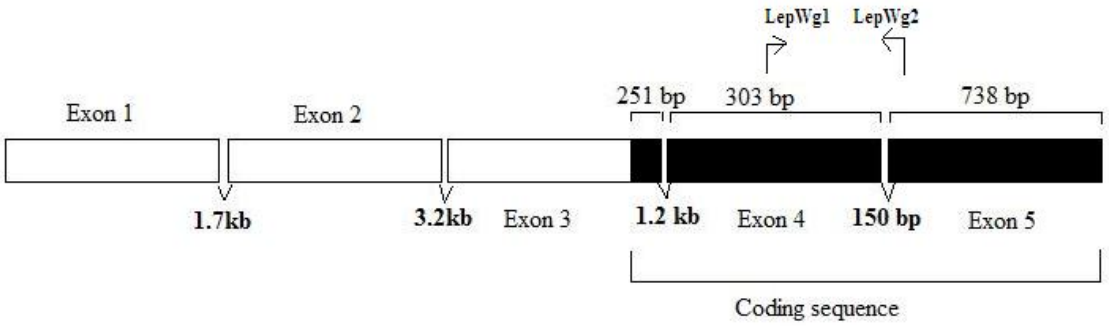

LepWg1
LepWg2
251 bp
303 bp
738 bp
Exon 1
Exon 2
1.7kb
3.2kb
Exon 3
1.2 kb
Exon 4
150 bp
Exon 5
Coding sequence

Figure 1.

Figure 2.

Figure 3.

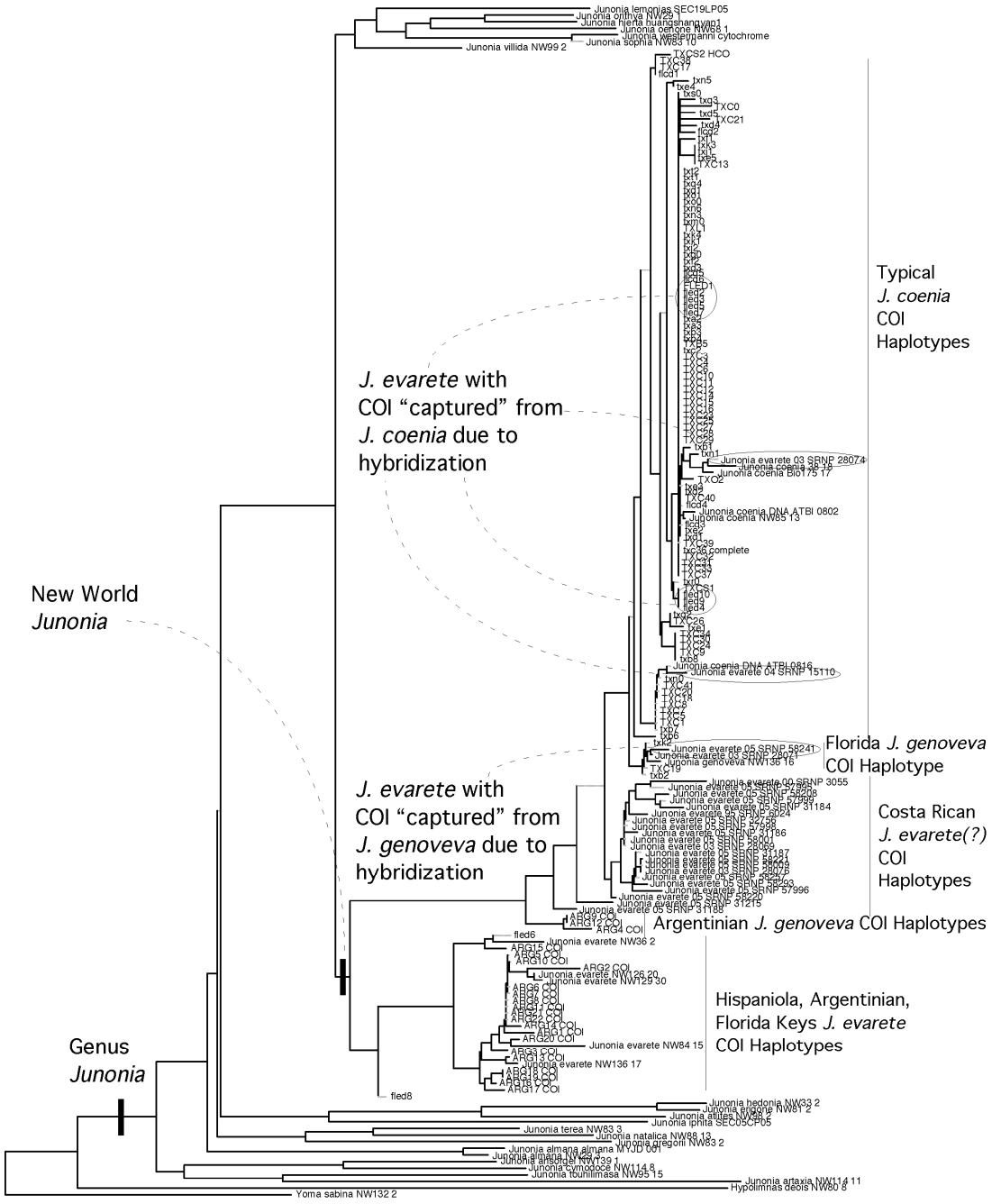

Typical
J. coenia
COI
Haplotypes
J. evarete with
COI "captured" from
J. coenia due to
hybridization
New World
Junonia
J. evarete with
COI "captured" from
J. genoveva due to
hybridization
Florida J. genoveva
COI Haplotype
Costa Rican
J. evarete(?)
COI
Haplotypes
Argentinian J. genoveva COI Haplotypes
Hispaniola, Argentinian,
Florida Keys J. evarete
COI Haplotypes
Genus
Junonia
– 0.001 substitutions/site

Figure 4.

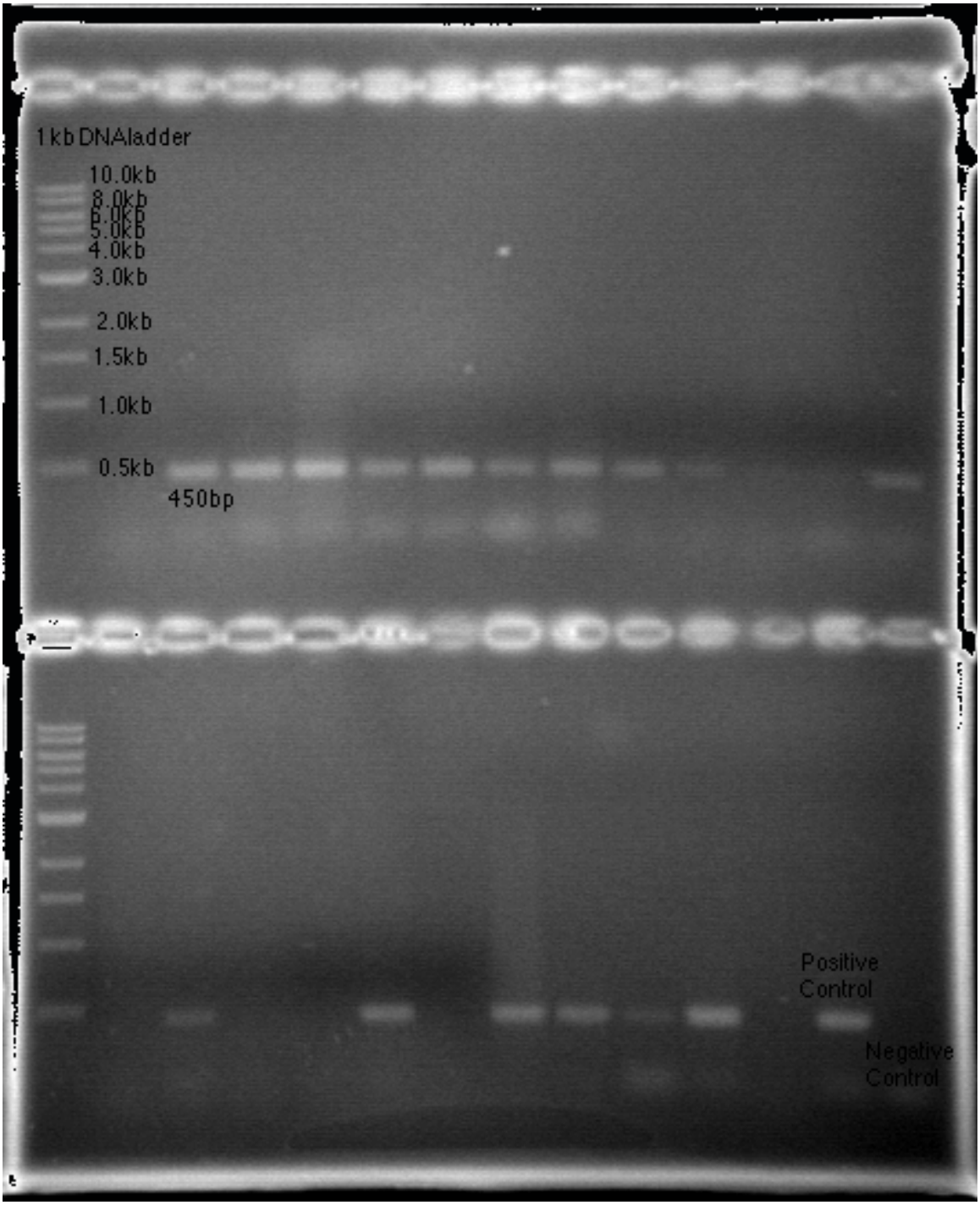
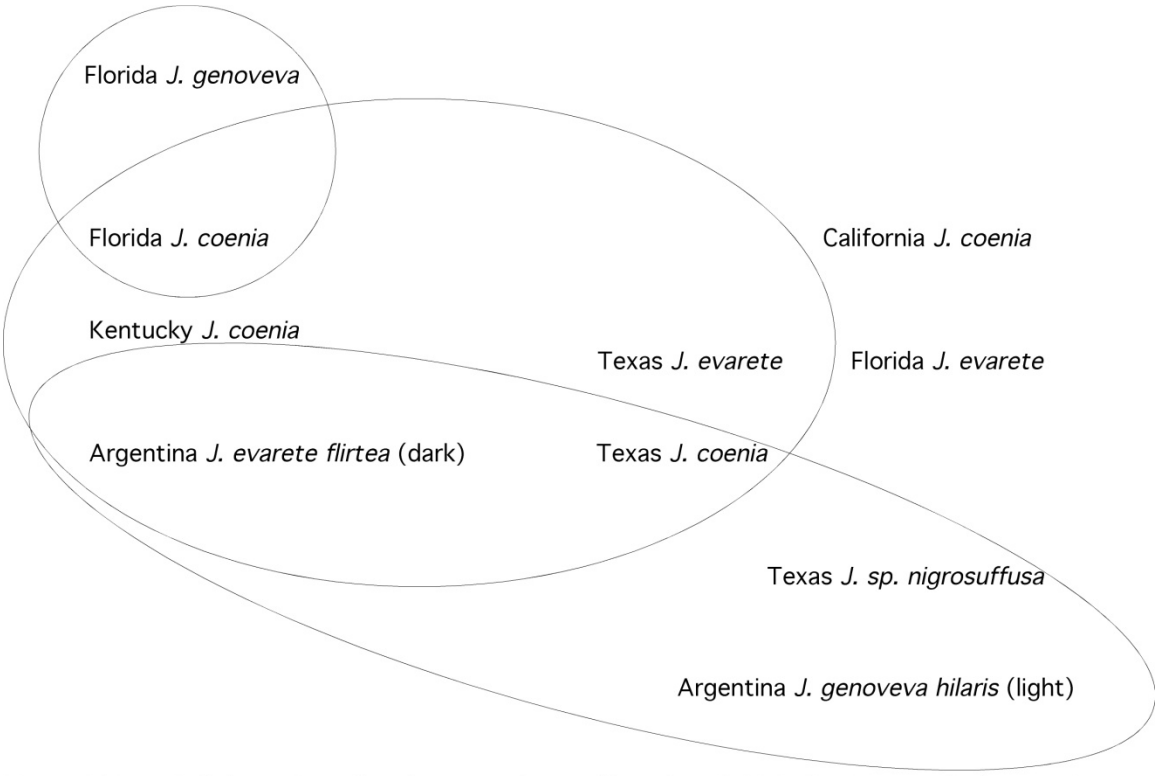

1kb DNAladder
10.0kb
8.0kb
6.0kb
5.0kb
4.0kb
3.0kb
2.0kb
1.5kb
1.0kb
0.5kb
450bp
Positive Control
Negative Control

Figure 5.



Taxa within a circle have Fst values between them of less than 0.01 indicating very little genetic differentiation.

Figure 6.